Genome Project and Genomic Approaches to Disease

In considering the Genome Project the following should be appreciated:

* **Important to understand enabling developments that historically had to be put in place.**

* **To grasp that the Human Genome Project is the product of a number of genome projects that interact.**

Key stages in the evolution of the Human Genome Project:

1956 Chromosomology.

The correct number of human chromosomes established.
The karyotype = the clinical geneticist's organ of investigation.

1966 Somatic Cell Genetics.

Mapping of genes for inborn errors of metabolism and cancer to individual chromosomes and regions.

1976 Mammalian Molecular Genetics.

Enabling cloning technology.

1986 Transgenic/Homologous Recombination Technology.

The ability to study gene function.

1996 Large Scale Genome Analysis - Genomics. Building genetic maps, physical maps and DNA sequencing.

What is meant by the Genome Project?

Aims

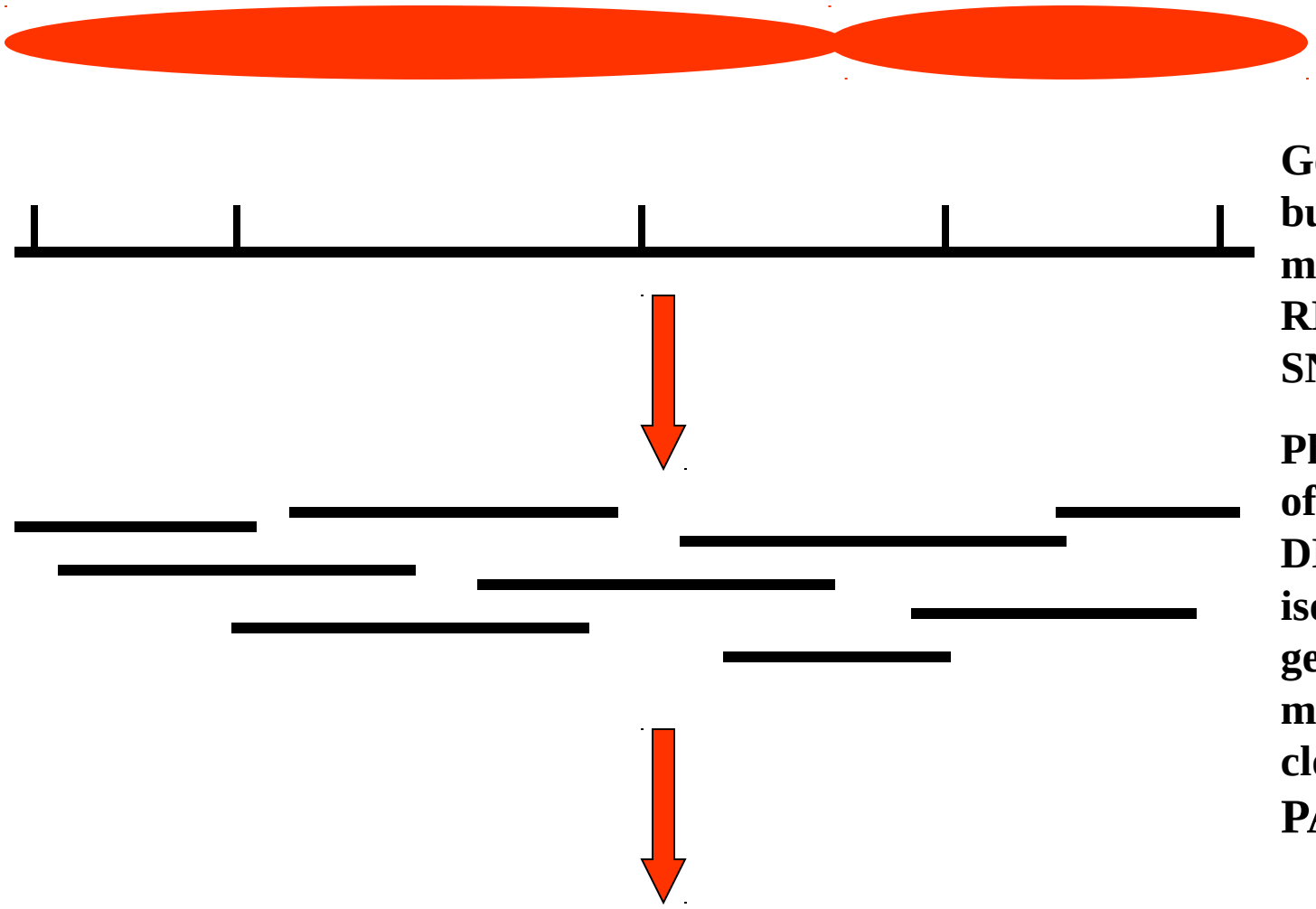
* Build genetic map of genome.

* Sequence genome and identify all genes.

* Understand function and biology of genes

* Map disease gene loci and identify disease genes and understand disease gene biology and related pathophysiology.

Strategy Driving the Genome Project - I



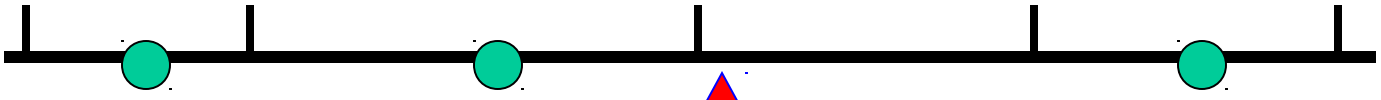
Genetic map built with poly-morphic markers: RFLP, VNTR, MS, SNP

Physical map of overlapping DNA clones isolated from genomic libraries made in large insert cloning vectors. PACs, BACs, YACs

Genome sequence based on ordered physical clone map

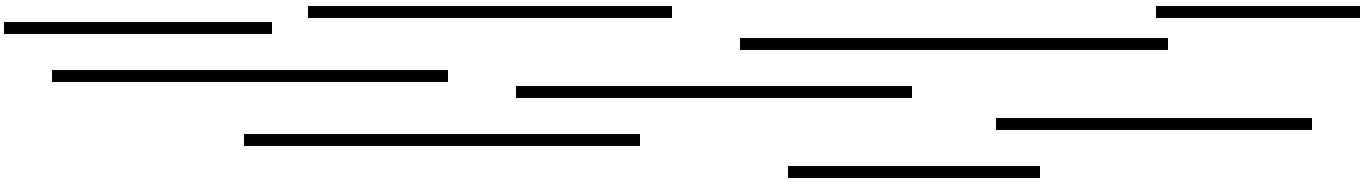
Computational gene identification

Strategy Driving the Genome Project - II



Genetic map - framework for family linkage studies to map disease gene loci

Gene Map linked to disease map: disease gene candidates



Physical map of overlapping DNA clones and associated sequence

Map transcripts on to clone and sequence map. Correlate to disease map.

cDNA libraries prepared from many tissues

Partial and total sequence from all gene transcripts (ESTs)

Database of partial and total mRNA sequences

Why study other genomes?

* Yeast

* C. elegans

*Drosophila

*Fugu (Puffer fish)

*Mouse

Test bed for the development of genomic analysis technologies.

Well developed genetic analysis with defined mutant phenotypes.

Simpler eukaryotic genomes; many genes shared with human.

Amenable to experimental manipulation. Elucidation of gene biological and biochemical function.

Why study other genomes?

*** Yeast**

*** C. elegans**

***Drosophila**

***Fugu (Puffer fish)**

***Mouse**

Genome 10x less complex than human genome.

(smaller introns)

Why study other genomes?

* Yeast

* C. elegans

*Drosophila

*Fugu (Puffer fish)

*Mouse

Model mammalian genome.

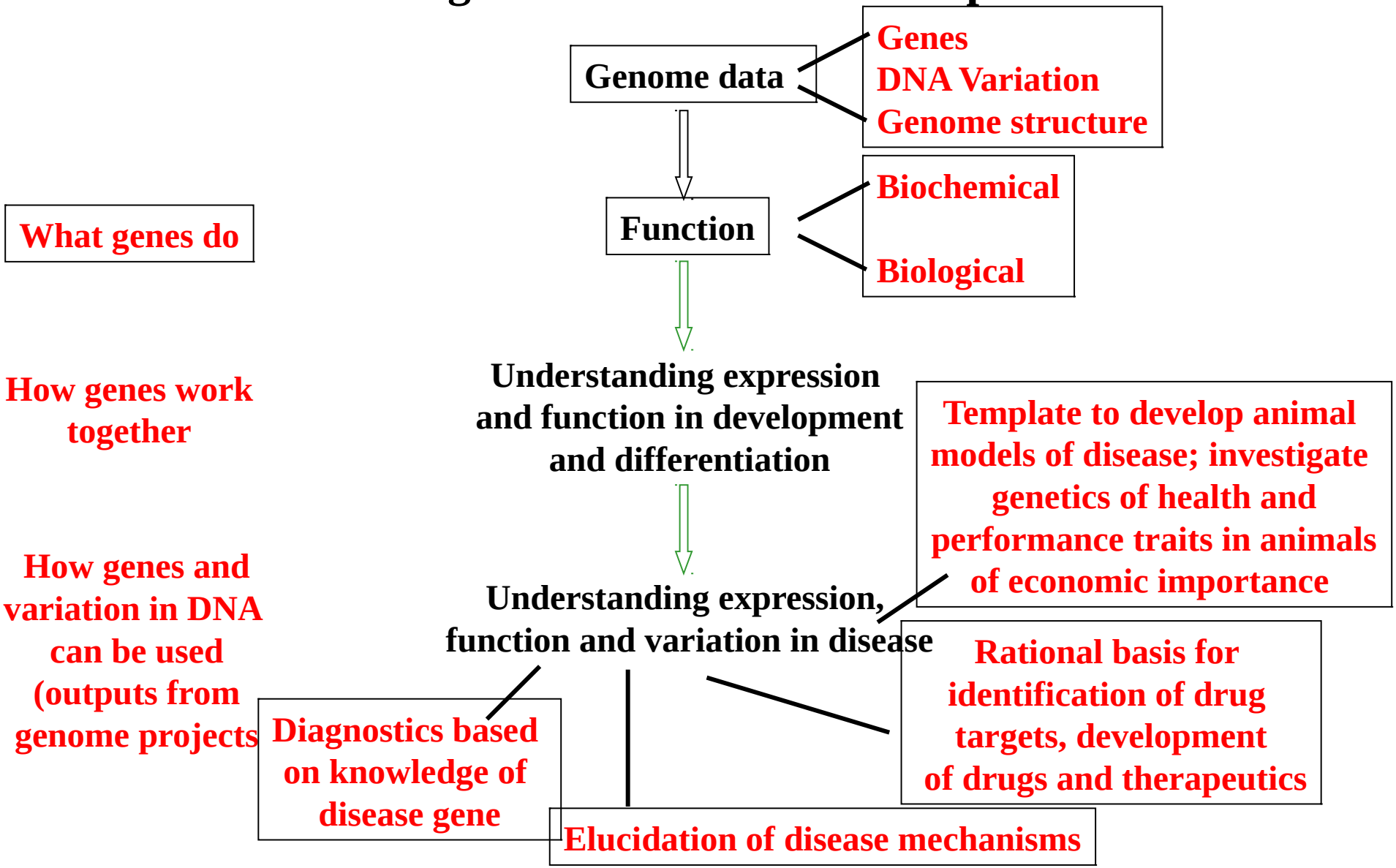
Sophisticated genetic map with defined phenotypes.

Comparative mapping allows prediction of human disease gene location for diseases modelled in mouse: by means of conserved gene order.

Transgenesis and homologous recombination permits investigation of gene function. Modelling of disease and therapeutic strategies.

Completion of genome sequence projects marks the start of genome-based biology

How can genome information be exploited ?



Genome Information and Resources Enable Genomic Approaches to Disease:

- **Candidate Gene Approaches**

- Association Studies Exploiting DNA Variation

- Functional Analysis

  - DNA Microarrays- genome wide expression analysis of genes - Transcriptome
  - Proteomics - defining proteins, their functions and interactions - Proteome
  - Gene Inactivation - defining function through gene silencing
  - Phenomics and Animal Models of Disease -Phenotype analysis – defining function through gene mutagenesis - Phenome

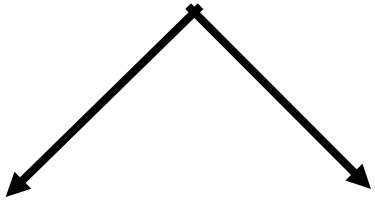
Two Variants:

* The Candidate Gene Approach

* The Positional Candidate Gene Approach

The Candidate Gene Approach

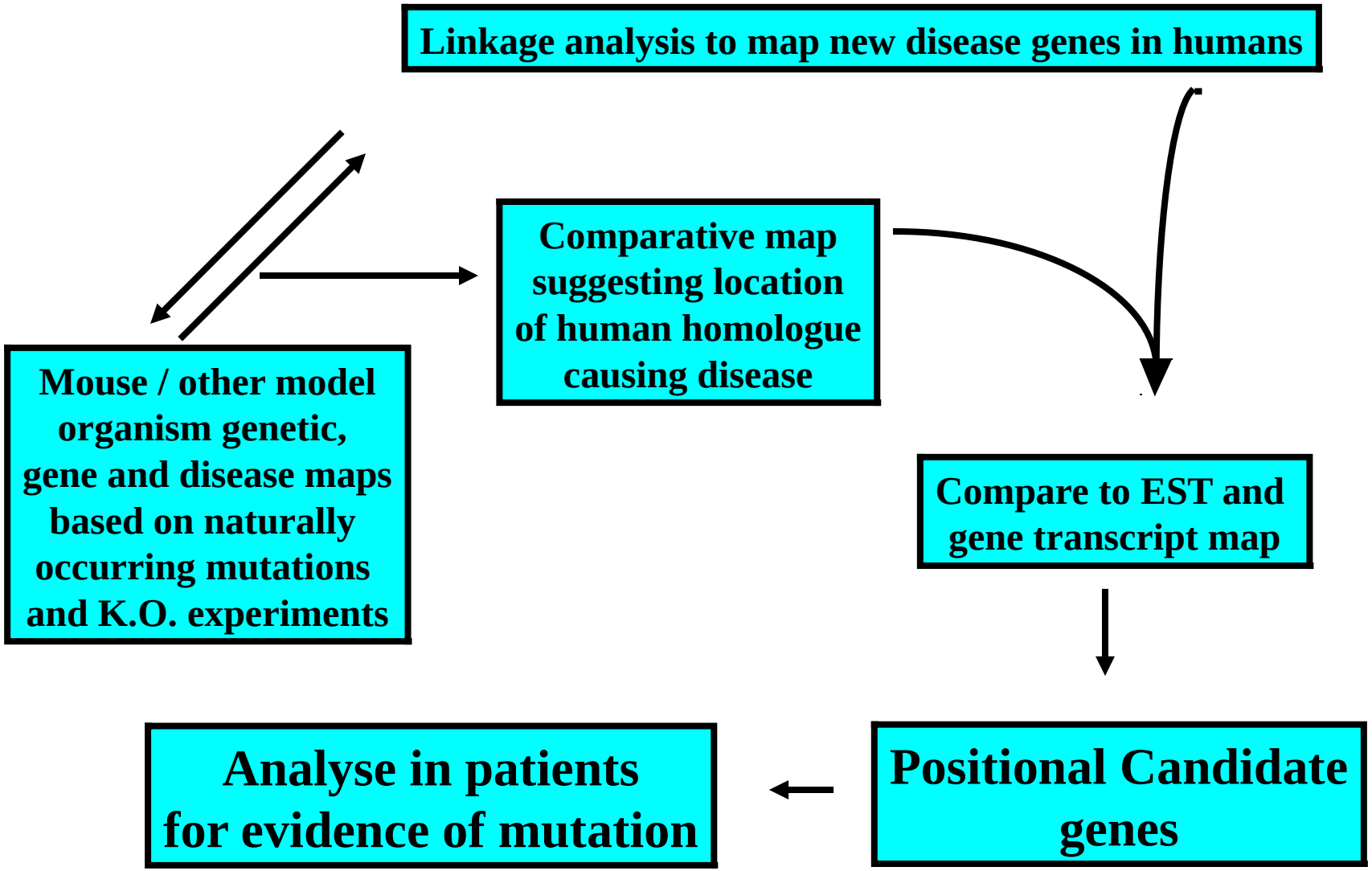
From an understanding of the biology of the disease, identify protein with appropriate function



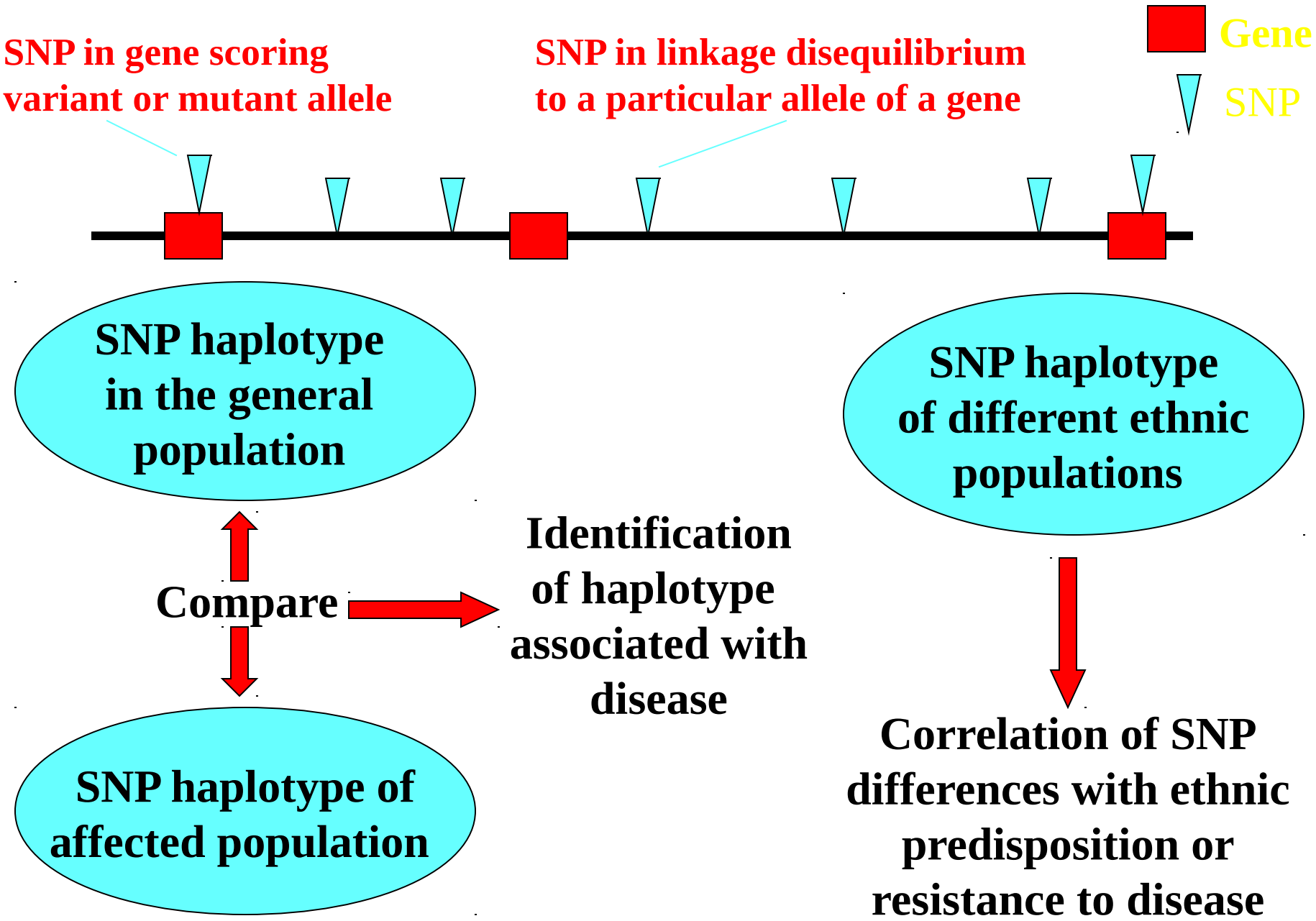
Correlate physical map location of gene encoding protein with any genetic location identified for disease by linkage

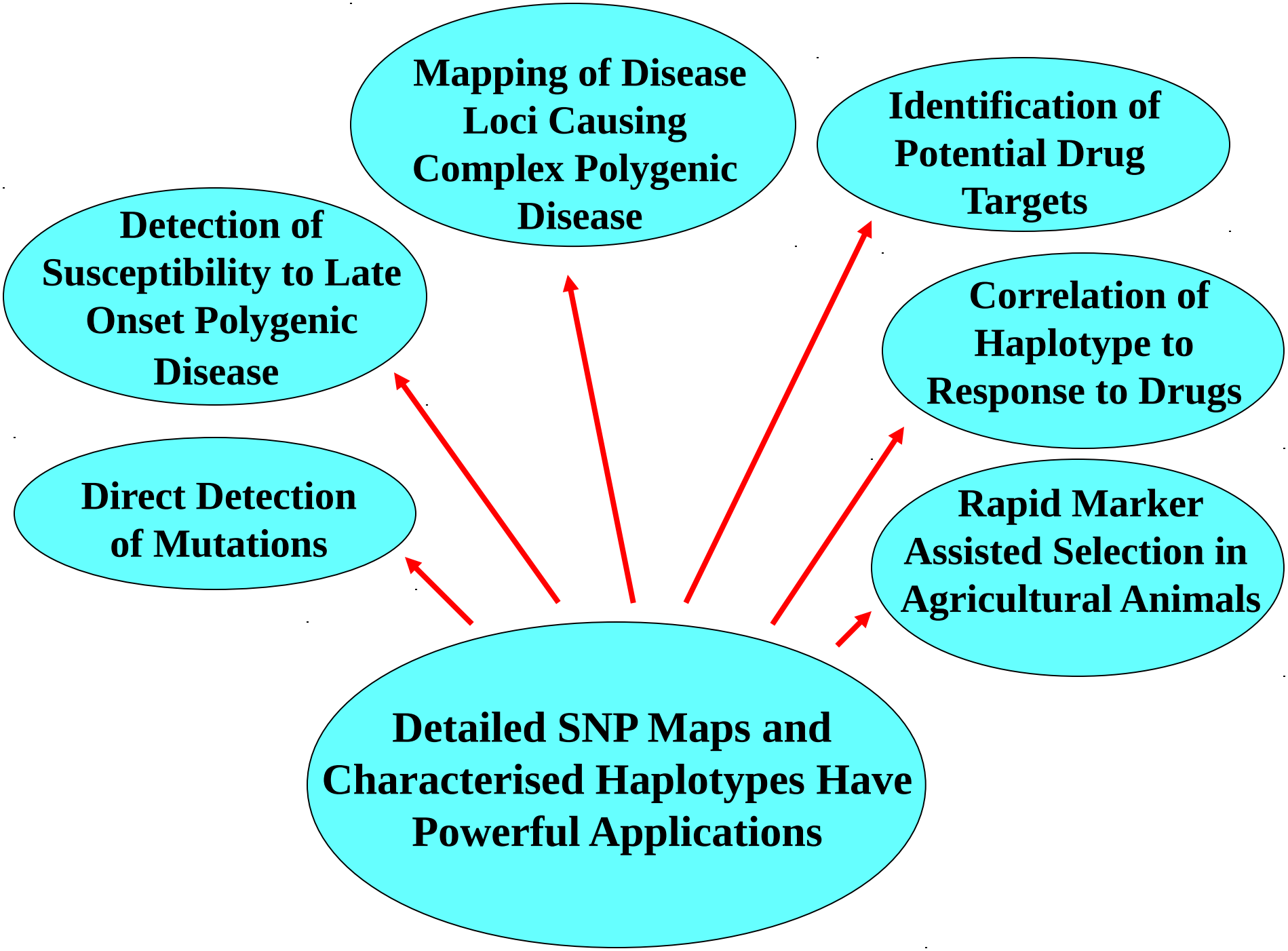
Mutation analysis of gene in patients to provide genetic evidence for involvement in the disease

The Positional Candidate Gene Approach



The Power of an SNP Map of the Human Genome




Mapping of Disease Loci Causing Complex Polygenic Disease
Identification of Potential Drug Targets
Detection of Susceptibility to Late Onset Polygenic Disease
Correlation of Haplotype to Response to Drugs
Direct Detection of Mutations
Rapid Marker Assisted Selection in Agricultural Animals
Detailed SNP Maps and Characterised Haplotypes Have Powerful Applications

Oligo tiling path across interval

Capture of sequences covering interval from DNA of affected individuals

Deep sequencing of entire interval to identify sequence Changes

Exon

Intron

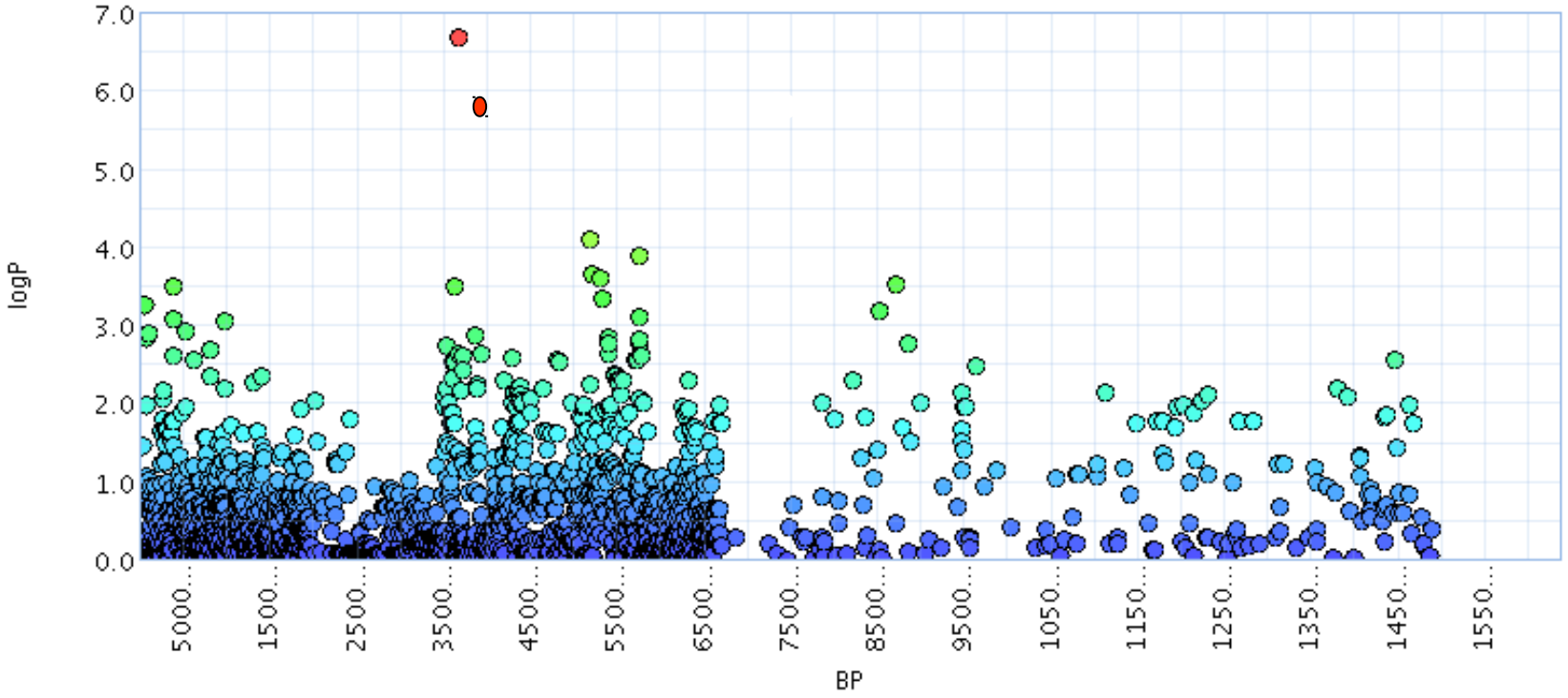
Promoter

Intergenic

Non-coding RNA

36000000

57000000



Relationship between Functional Approaches

Differential gene expression defining disease processes / state

MICROARRAYS

Differential protein expression and activity defining disease processes / state

PROTEOMICS

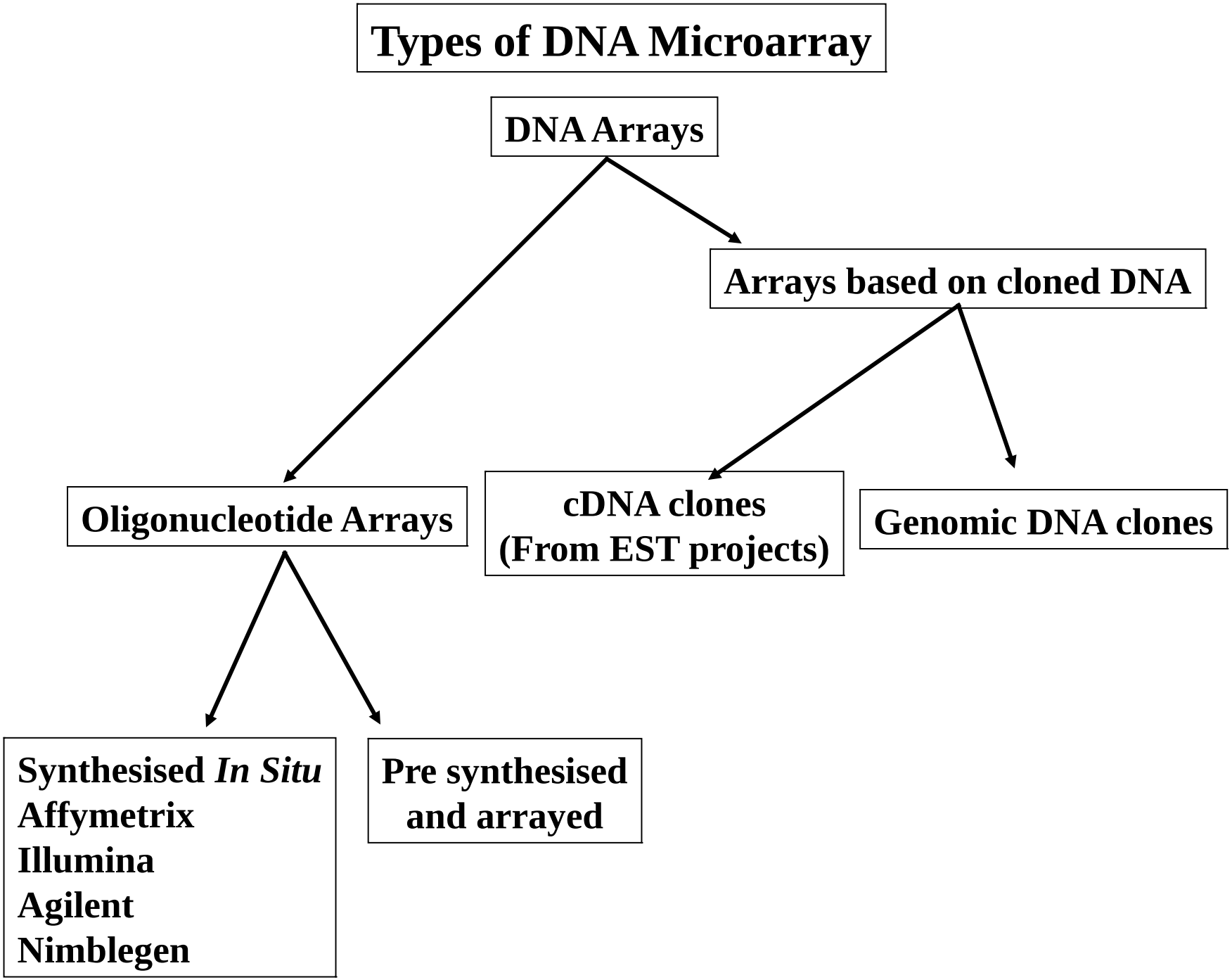
Transcriptional events and mRNA stability

Post-transcriptional events: Translational control
Protein turnover
Protein modification

Candidate genes for functional studies using gene silencing modelled in mouse etc

Genes leading to disease not defined by transcription profiling or proteome analysis

MUTAGENESIS / PHENOTYPE SCREEN

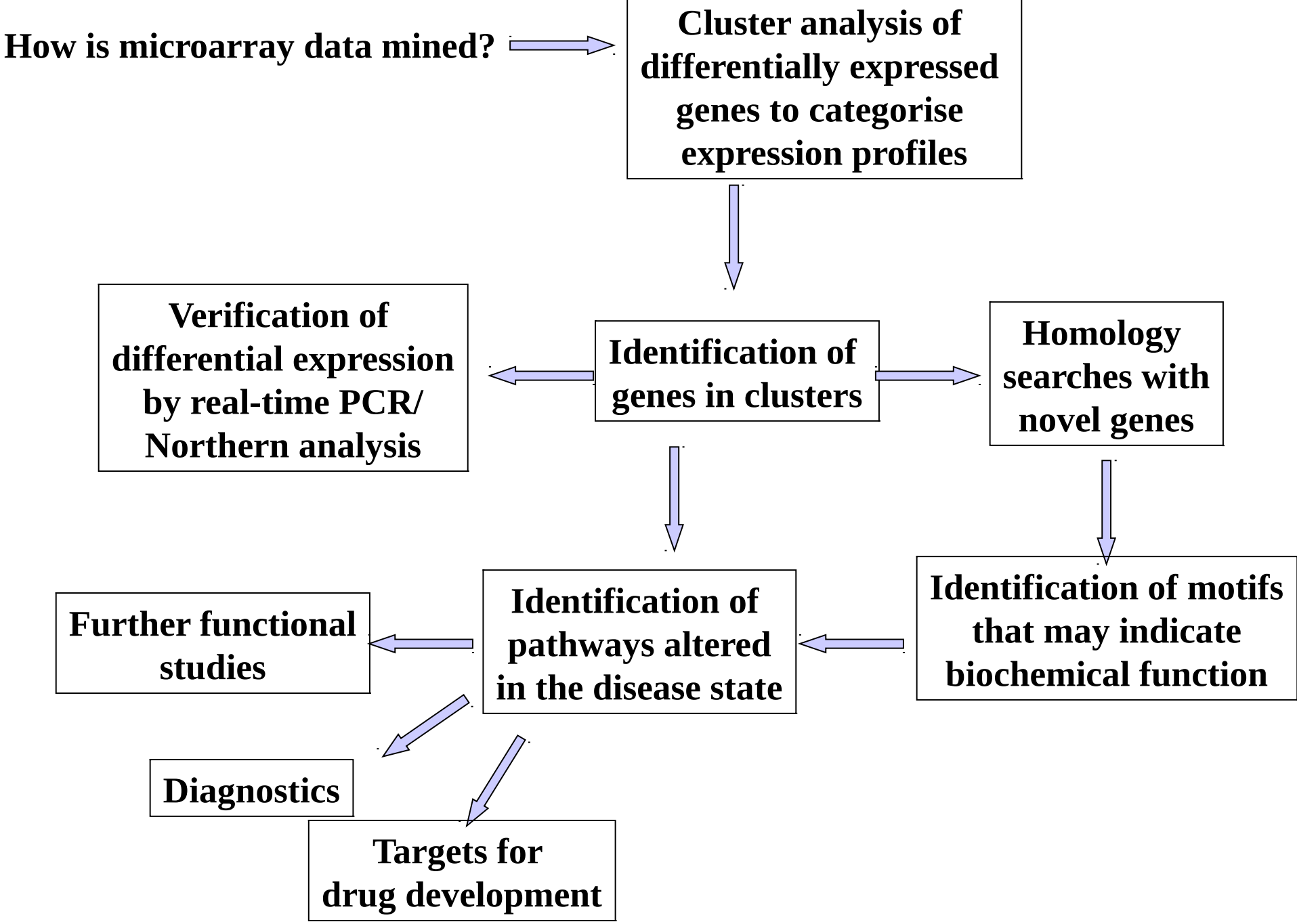

Types of DNA Microarray
DNA Arrays
Arrays based on cloned DNA
Oligonucleotide Arrays
cDNA clones
(From EST projects)
Genomic DNA clones
Synthesised *In Situ*
Affymetrix
Illumina
Agilent
Nimblegen
Pre synthesised
and arrayed

Why is transcription profiling so powerful?

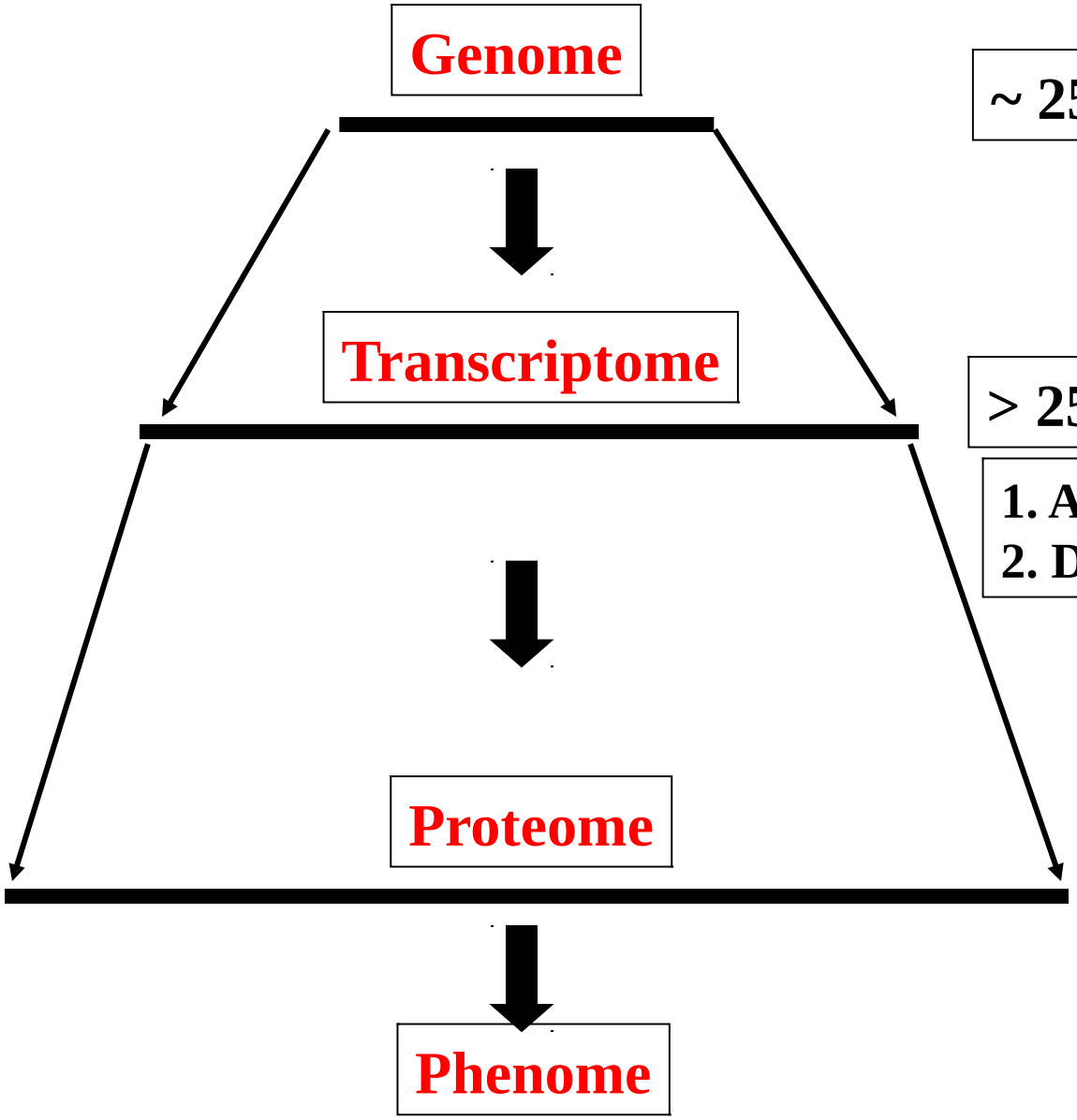
***Defines genetic pathways and temporal transcription patterns in normal differentiation and disease processes.**

***Defines genes that show differential expression as potential targets for development of new drugs - Pharmacogenomics.**

***Transcriptional signatures of disease states can function to diagnose disease sub-types and indicate prognosis and most effective treatment eg leukaemias caused by translocations.**


How is microarray data mined?
Cluster analysis of differentially expressed genes to categorise expression profiles
Verification of differential expression by real-time PCR/ Northern analysis
Identification of genes in clusters
Homology searches with novel genes
Identification of motifs that may indicate biochemical function
Identification of pathways altered in the disease state
Further functional studies
Diagnostics
Targets for drug development

The Challenge of Proteomics: there are many more proteins than than genes.



~ 25,000 genes

> 25,000 transcripts

1. Alternative promoters.
2. Differential splicing.

>> 25,000 proteins

1. Protein isoforms
2. Post-translational modifications.

How is the proteome being studied?:

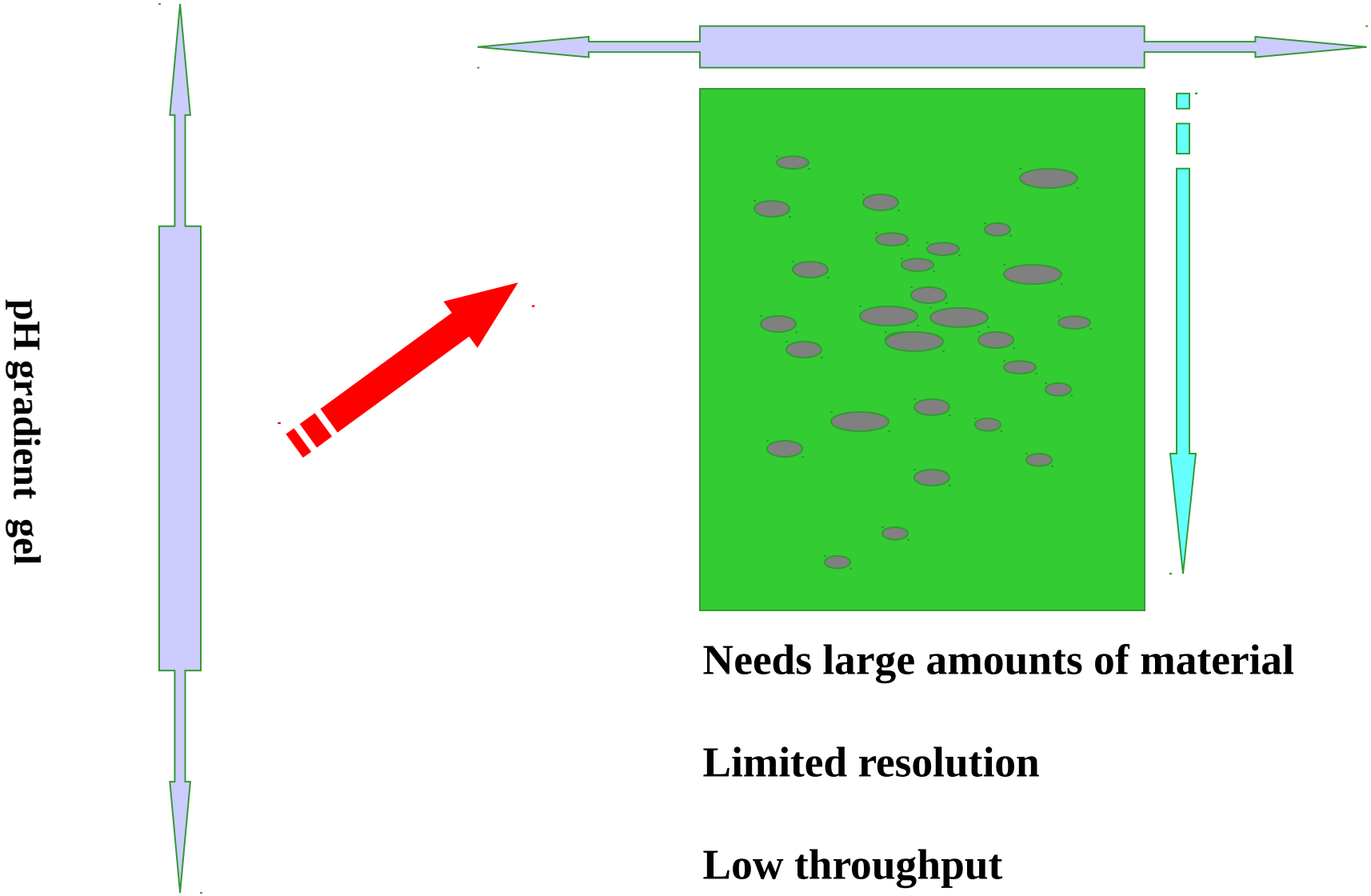
***2D Gels: Can resolve proteins that have undergone modifications. Phosphorylation, glycosylation etc.**

***Mass Spectroscopy: resolves oligopeptide patterns to identify proteins. Can be combined with 2D gels.**

***Protein Arrays: Analogous to DNA microarrays. Allows large scale protein / antibody interactions (useful in diagnostics). Allows protein and protein DNA interactions to be studied.**

First Dimension
Iso-electric focusing

Second Dimension
SDS gel - separation on MW



Proteins focus at their charge in the pH gradient

Needs large amounts of material

Limited resolution

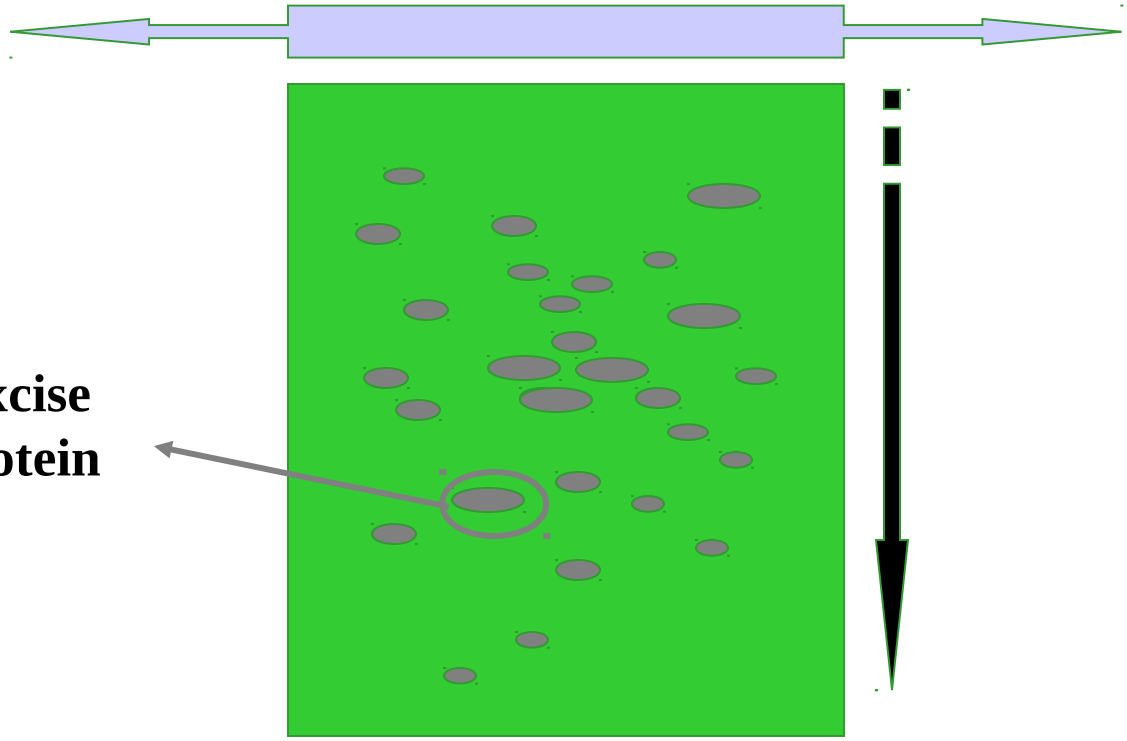
Low throughput

Digest into Oligopeptides

Mass Spec analysis to identify protein

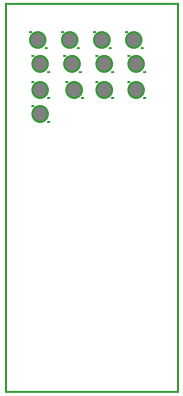
Characterise protein

Excise protein



Applications of protein based arrays

Recognition Molecule:
Arrayed on chip



Chemical - hydrophobic
hydrophilic
Anionic / Cationic
Metal ion

Biological - Antibody
Receptor
DNA / RNA

Selection of proteins based on different physical properties

Selection of proteins based on protein-protein and protein-DNA / RNA interactions

Mass Spec analysis of selected proteins

*** High throughput identification of proteins and complexes**

*** High throughput isolation of interacting proteins compared to Yeast Two Hybrid system**

*** Diagnostics based on antibody / antigen interactions e.g. ovarian cancer markers in serum**

Linking Gene to Phenotype in Model Organisms

*** Gene silencing through homologous recombination - genotype driven screen**

- direct inactivation

- conditional inactivation by means of site-specific recombination using Cre recombinase

Labour intensive - cannot be applied on a genomic scale
Does not always cause a phenotype - redundant genetic pathways

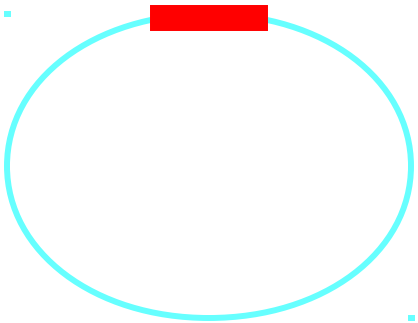
*** Gene silencing through the use of Interference RNA - RNAi**

Highly specific distinguishing gene family members
Rapid and can be applied on a genomic scale e.g. plants, C. elegans, Drosophila

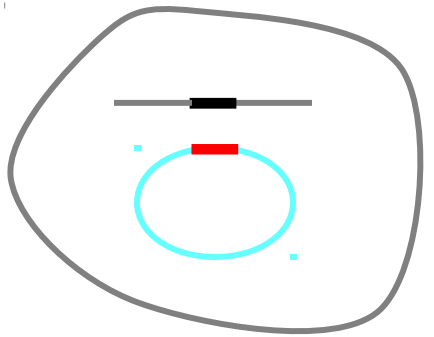
*** ENU mutagenesis and phenotype driven analysis**

Identifies genes by mutation saturation
Introduces different mutations per gene thus permitting analysis of functional domains

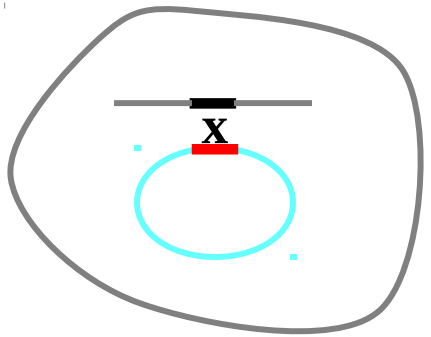
Preparation of targeting construct containing part of gene



Transfection into Embryonal Stem Cells (ES)

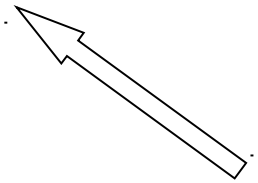


Recombination with endogenous cellular gene

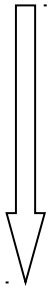


ES cell with inactivated gene

Introduction of ES cells into developing mouse blastocycts

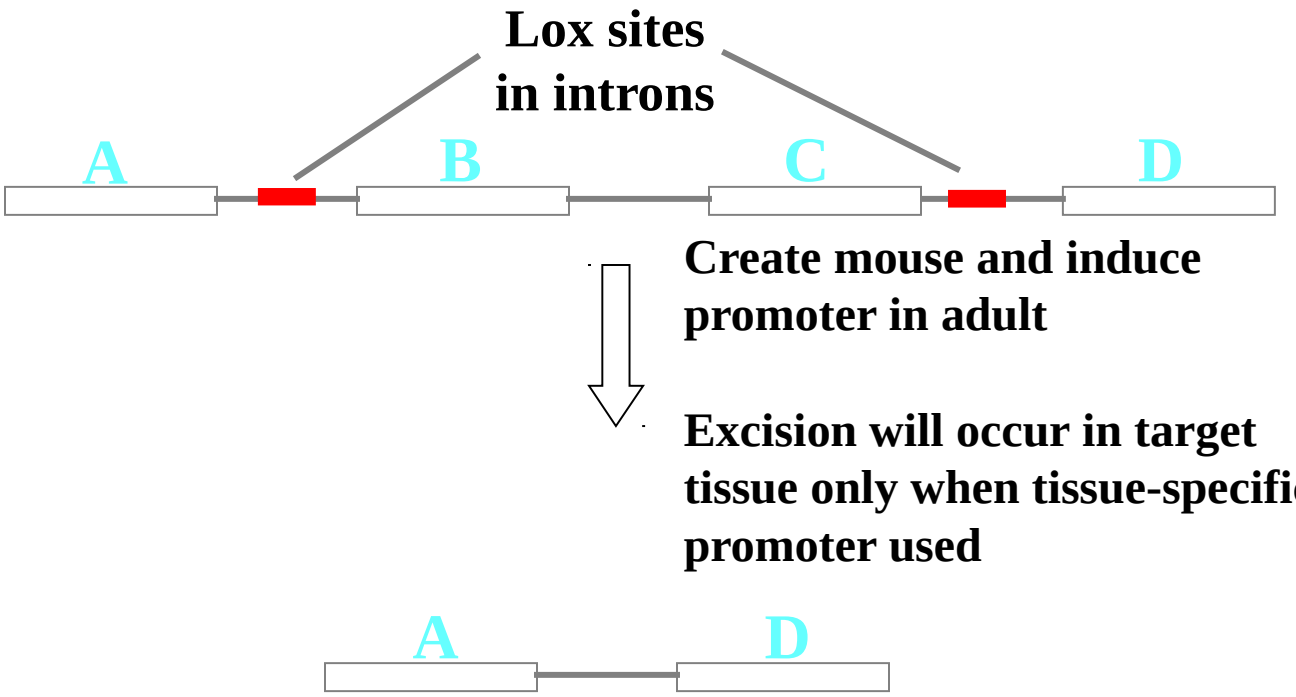


Embryo develops into mouse with specific gene inactivated and associated phenotype



Needed when direct inactivation results in an embryonic lethal phenotype

Instead of directly silencing the gene in the ES cell, exons of the gene are flanked by Lox sites that direct site -specific recombination by the Cre recombinase. This is done in an ES cell transgenic for the Cre gene under the control of an inducible or tissue-specific promoter.

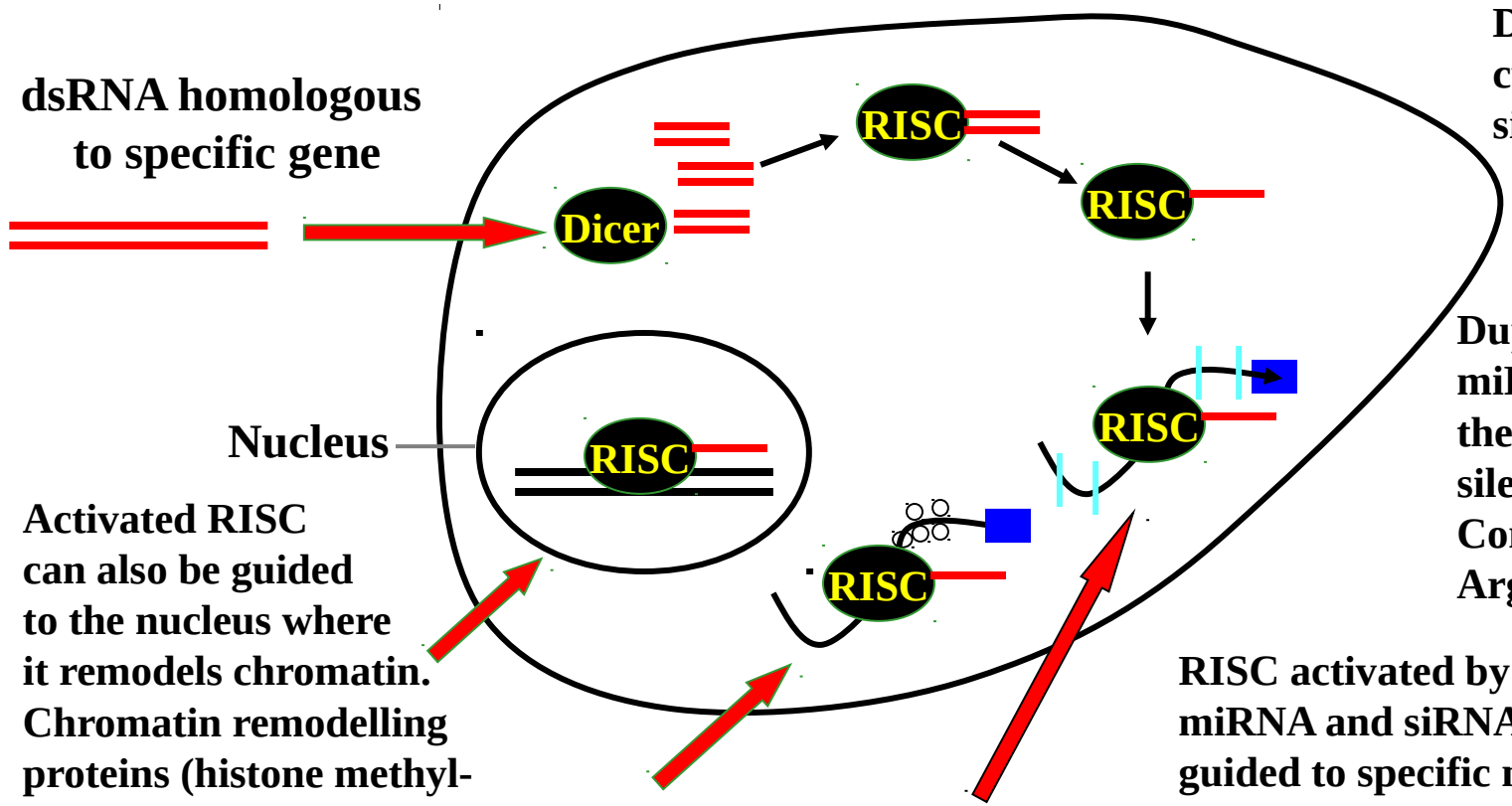


***Developed and applied initially in plants, C. elegans and Drosophila and to mammalian cells in tissue culture**

***Highly gene specific capable of distinguishing related genes**

***Functions by subverting endogenous cellular pathways that utilise small double-stranded RNA molecules involved in gene regulation**

Biogenesis of siRNA and miRNA



Dicer complex cuts dsRNA into siRNA and miRNA

Duplex siRNA and miRNA is targeted to the RISC (RNA induced silencing complex). Contains member of the Argonaute family.

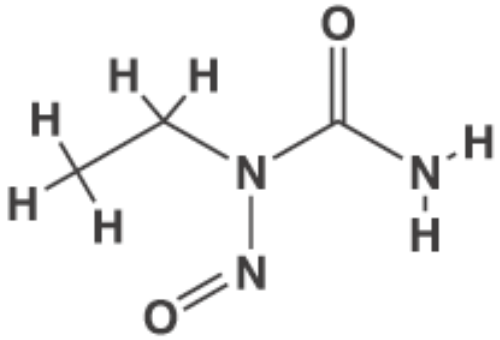
RISC activated by rendering miRNA and siRNA single-stranded and guided to specific mRNA by homology

Activated RISC can also be guided to the nucleus where it remodels chromatin. Chromatin remodelling proteins (histone methyl-transferases, acetyl transferases, HDACs, and ubiquitinating enzymes bind RNA)

Activated RISC can inhibit mRNA translation (target 3'UTR – imperfect match)

Activated RISC can degrade mRNA where there is a perfect match; Rnase H-like activity in piwi domain of Argonaute

N – ethyl – N – nitrosourea
ENU
An alkylating agent that transfers ethyl group to nucleobases



Targets spermatogonial stem cells and induces single base changes every 1 – 2 Mb at a rate of 1 per 700 gametes
A>T base transversions; AT>GC transitions; also (less frequently) GC>AT transitions

***ENU induces loss of function, gain of function, altered function mutations - wider range of phenotypes depending upon region of gene that is mutated**

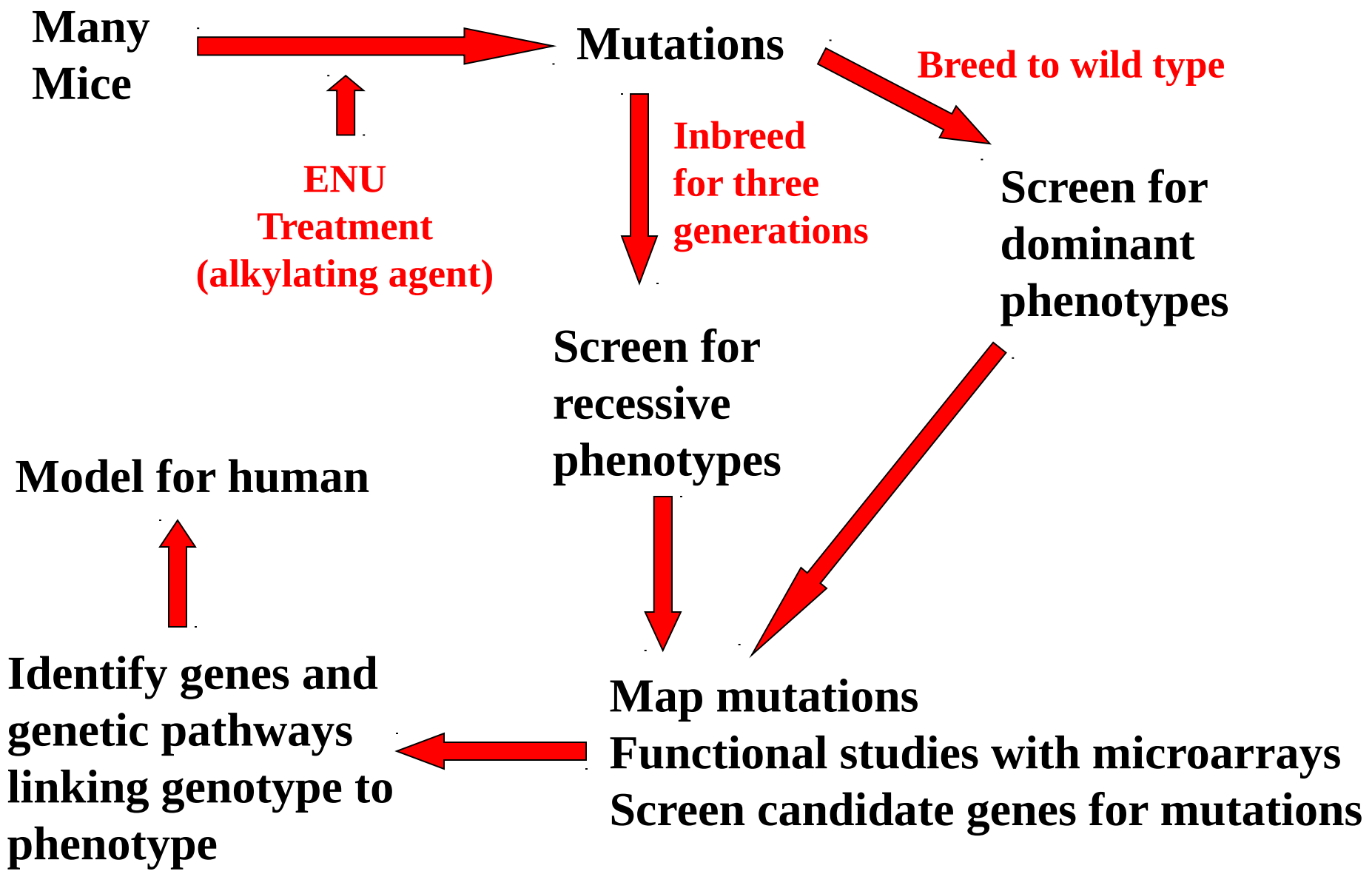
***ENU mutagenesis is not biased - K.O. and RNAi *a priori* define the gene**

***ENU can alter more than one gene**

***Creates complex phenotypes**

***Phenotpyes can be generated rapidly**

An Approach to the Phenome: Rapid Derivation of function from phenotype. cf Transgenic / K.O. mice



Summary

1. **Genome projects have produced genome information resources that allow rapid identification of disease genes e.g. Maps, comparative maps, DNA sequence, SNPs**

2. **Genome projects have produced resources for functional analysis e.g. microarray resources, clone banks**

3. **Genome projects have forced the development of technologies to approach biology on a genome wide scale e.g. transcription profiling, proteomics and gene inactivation strategies**

Summary

1. **Genome projects have produced genome information resources that allow rapid identification of disease genes e.g. Maps, comparative maps, DNA sequence, SNPs**

2. **Genome projects have produced resources for functional analysis e.g. microarray resources, clone banks**

3. **Genome projects have forced the development of technologies to approach biology on a genome wide scale e.g. transcription profiling, proteomics, gene inactivation strategies and massively parallel sequencing approaches**

Strachan and Reid Human Molecular Genetics 4th Edition (2011)

Chapters 12 and 16